Econometric Analysis

Ch. 5 Instrumental Variables Estimation of Single-Equation Linear Models

Ryuichi Tanaka

5.1 IV and Two-stage Least Squares
5.1.1 Motivation for IV Estimation

The population model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u$

$(y, \mathbf{x})$: observable random vector
u : unobservable random disturbance
$\boldsymbol{\beta}$: parameters to be estimated

Assumptions for the consistency of OLS estimator

(1) $E(u) = 0$ (automatically satisfied if the model has a constant term)
(2) $Cov(x_j, u) = 0,\ j = 1, 2, \cdots, K.$ (xs are exogenous)

Instrumental Variable Method

Ex.
The population model: $\log(wage) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 educ + u$

u may contain unobservables such as ability and thus correlate to *educ*
＝＞OLS estimator is not consistent

=> IV method as a general approach to endogeneity problem

Definition of Instrumental Variables

The population model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u$
Assumption: $E(u) = 0, Cov(x_j, u) = 0,\ j = 1, 2, \cdots, K-1, Cov(x_K, u) \neq 0$

Instrumental variable(z_1) :
(1) $Cov(z_1, u) = 0$ (z_1 is exogenous)
(2) In the linear projection: $x_K = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + r_K$,
$\theta_1 \neq 0$ (z_1 is correlated with x_K)

z_1 is partially correlated with x_K once the other exogenous variables have been netted out

Instrumental variables: $(x_1, \ldots, x_{K-1}, z_1)$
$(x_1, \ldots, x_{K-1})$ are instruments to their own

Consistency

LP $x_K = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + r_K$:
a reduced form equation for x_K

Plugging the reduced form in the population model, we obtain a reduced form of y:
$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_{K-1} x_{K-1} + \lambda_1 z_1 + v, \quad v = u + \beta_K r_K$
$\alpha_j = \beta_j + \beta_K \delta_j, \quad \lambda_1 = \beta_K \theta$

Since $v = u + \beta_K r_K$ is uncorrelated with $(x_1, \cdots, x_{K-1}, z_1)$, we can estimate (α_j, λ_1) consistently by OLS

However, what we want to know is not (α_j, λ_1) but β in the population model
=> Can we identify the βs?

Identification of Structural Parameters

Identification: β is identified if β can be written as a population moment of observables

$y = x\beta + u, x = (1, x_2, \cdots, x_K), z = (1, x_2, \cdots, x_{K-1}, z_1)$ 1xK vec. of all exo. var. ($E(z'u) = 0$)

$E(z'y) = [E(z'x)]\beta + E(z'u) = [E(z'x)]\beta$ (K equations with K unknowns)

If KxK matrix $E(z'x)$ has full-rank, $\beta = [E(z'x)]^{-1} E(z'y)$

If $E(x'x)$ has full-rank, $E(z'x)$ has full-rank if and only if $\theta_1 \neq 0$
The assumption $\theta_1 \neq 0$ is crucial for identification

Example 5.1

The population model: $\log(wage) = \beta_0 + \beta_1 exper + \beta_2 exper^2 + \beta_3 educ + u$
u is correlated with $educ$ => Find IV for $educ$!

Candidate 1 mother's education: $\theta_1 \neq 0$, but $Cov(z_1, u) = 0$ may not be satisfied
Candidate 2 last digit of social security number: $Cov(z_1, u) = 0$, but $\theta_1 \neq 0$ not satisfied
Candidate 3 $z_1 = 1$ if born in the first quarter: $Cov(z_1, u) = 0$ and $\theta_1 \neq 0$
(Angrist and Krueger 1991, Q.J.E.)

Why $\theta_1 \neq 0$?
American primary schools start in September
Children whose age is 6 at january 1st of the year are allowed to enrol in the first grade
School attendance is compulsory until 16th birthday
Children born in the first quarter face shorter period of compulsory education

Numerical example of 5.1

Compare 2 persons

1:Born in December, 1980, 2: Born in January, 1981
6 years old at Jan. 1st: 1987, 1988

Enroll in the 1st grade: September 1987, September1988

End of compulsory education (16th birthday): Dec. 1996, Jan. 1997

Person 2 faces shorter period of compulsory education than the person 1

How to find a good IV: Randomization

Randomization (random selection for the program):
Individuals are randomly selected to be eligible for a program

Ex. Job Training Probram(JTP): $\log(wage) = \beta_0 + \beta_1 JTP + \mathbf{x}\boldsymbol{\beta}_2 + u$
If u contains unobserved ability and if able person tends to participate in the JTP, JTP is endogenous

If the eligibility for the JTP is randomly distributed, eligibility is exogenous with no correlation with ability. Moreover, since eligibility is needed for the participation to the JTP, $\theta_1 \neq 0$

Natural Experiment 1 : Exgogenous change in institution and environment

Ex. Acemoglu, Autor and Lyle (2004 JPE): relation between female labor supply and wage
Female labor supply correlates with ability (endogenous)

Exogenous increase in female labor supply due to mobilization of male workers during the WWII
=> state level mobilization rate as an IV for female labor supply

Other examples :
Exogenous change in compulsory education system as IV for education

Natural Experiment 2: geographic information

Hoxby(1994) : Estimate the effect of school choice on school productivity (academic performance, test score)
Measure of school choice: the number of school districts in metropolitan area

# of school districts is endogenous (negative correlation): if performance of one district is high, people move in the district and other districts will merge to the district

# of school districts is positively correlated with the # of rivers in the area (# of rivers is IV for the # of school districts)

Exogenous trait

Evans and Schwab(1995): Estimate the academic performance oif students in Catholic high schools
Attendance to a Catholic high school (endogenous)
IV: if the student is Catholic (exogenous)

5.1.2 Multiple Instruments: Two-stage Least Squares

Structural Model : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u$

$E(u) = 0, Cov(x_j, u) = 0,\ j = 1,2,\cdots,K-1, Cov(x_K, u) \neq 0$

Assume that multiple IVs for x_K are available : $z = (1, x_2, \cdots, x_{K-1}, z_1, \cdots z_M)$

If we use z_j one by one as an instrument for x_K, we have M IV estimates

Moreover, any linear combination of $z = (1, x_2, \cdots, x_{K-1}, z_1, \cdots z_M)$ can be IV

$=>$ Which IV should we use?

$=>$ 2SLS is the most *efficient* IVestimator!

Two-stage Least Squares

Stage 1 :Estimate $x_K = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + \theta_2 z_2 + \cdots + \theta_M z_M + v$

by OLS, and construct the fitted value $\hat{x}_K = \hat{\delta}_0 + \hat{\delta}_1 x_1 + \hat{\delta}_2 x_2 + \cdots + \hat{\delta}_{K-1} x_{K-1} + \hat{\theta}_1 z_1 + \hat{\theta}_2 z_2 + \cdots + \hat{\theta}_M z_M$

Stage 2 : Estimate $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{K-1} x_{K-1} + \beta_K \hat{x}_K + u$ by OLS

Interpretation :in the 2SLS, we use IVwith the highest correlation with x_K

IVwith the highest correlation with x_K can be obtained as the LP of x_K on the other IVs:

$x_K = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + \theta_2 z_2 + \cdots + \theta_M z_M + r_K$

$x_K = x_K^* + r_K, Cov(x_K^*, u) = 0$ (the source of endogeneity of x_K is in r_K, $Cov(r_K, u) \neq 0$)

Two-stage Least Squares (cont.)

If we know the parameters in $x_K = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_{K-1} x_{K-1} + \theta_1 z_1 + \theta_2 z_2 + \cdots + \theta_M z_M + r_K$,
we use x_K^* as an IV, but we do not know these population parameter ex ante

However, since we can consistently estimate these parameters by OLS (variables in the first stage are exogenous), we use
$\hat{x}_K = \hat{\delta}_0 + \hat{\delta}_1 x_1 + \hat{\delta}_2 x_2 + \cdots + \hat{\delta}_{K-1} x_{K-1} + \hat{\theta}_1 z_1 + \hat{\theta}_2 z_2 + \cdots + \hat{\theta}_M z_M$
in the second stage estimation

Intuitively, we get rid of the endogeneity in the first stage

5.2 General Treatment of 2SLS
5.2.1 Consistency

The population model: $y = \mathbf{x}\boldsymbol{\beta} + u$

Assumption 2SLS.1: For some 1xL vector $\mathbf{z}, E(\mathbf{z}'u) = 0$ (IV are exogenous)

IV should contain all exogenous variables in the population model
and at least one exogenous variable excluded from the population model

$E(u \mid \mathbf{z}) = 0 \Rightarrow A2SLS.1$

5.2.1 Consistency (cont.)

Assumption 2SLS.2 : (a) rank$E(\mathbf{z}'\mathbf{z}) = L$; ($b$)rank$E(\mathbf{z}'\mathbf{x}) = K$ (rank condition)

Theorem 5.1 : (consistency of 2SLS)
Under A2SLS.1 and A2SLS.1, the 2SLS estimator obtained from a random sample is consistent for $\boldsymbol{\beta}$

rank condition is necessary for identification. Roughly, this condition is satisfied if z and x are correlated enough

Order condtion: $L \geq K$ (necesarry for rank condition)
we need at least same number of IVs as the explanatory variables

5.2.2 Asymptotic Normality of 2SLS

$N^{-1/2}\sum_{i=1}^{N}\mathbf{z_i}'u_i$ is asymptotically normal (by CLT)

=> $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically normal

A2SLS.3 (homosckedasticity): $E(u^2 z'z) = \sigma^2 E(z'z)$ where $\sigma^2 = E(u^2)$

$E(u^2 \mid z) = \sigma^2$ => A2SLS.3

Theorem 5.2 : Under A2SLS1, 2 and 3, $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically normally distributed with mean zero and variance matrix $\sigma^2\{E(\mathbf{x}'\mathbf{z})[E(\mathbf{z}'\mathbf{z})]^{-1}E(\mathbf{z}'\mathbf{x})\}^{-1}$.

Consistent estimator of asymptotic variance

Consistent estimator of σ^2: $\hat{\sigma}^2 = (N-K)^{-1}\sum_{i=1}^{N}\hat{u}_i^2$ where $\hat{u}_i = y_i - x_i\hat{\beta}$ (2SLS residuals)

Consistent estimator of asymptotic variance $\sigma^2\{E(x'z)[E(z'z)]^{-1}E(z'x)\}^{-1}$

$\hat{\sigma}^2(\sum_{i=1}^{N}\hat{x}_i'\hat{x}_i)^{-1}$

Notice :2SLS residuals ($\hat{u}_i = y_i - x_i\hat{\beta}$) is NOT the residuals from the second stage $(y_i - \hat{x}_i\hat{\beta})$!

5.2.3 Asymptotic Efficiency of 2SLS

<u>Theorem 5.3</u> (Relative efficiency of 2SLS): Under A2SLS1, 2 and 3, the 2SLS estimator is efficient in the class of all instrumental variables estimators using instruments linear in z.

<u>Implication</u> : the more IV we use, the more efficient in the limit
<u>Why?</u> When z (L>K) IVs are available, Using only a part of z (z') **as IV means that we are using only a special case of a** linear combination of Z as IV

<u>Note:</u> if we use too many Ivs, we may face a finite sample problem (what if we use "IV" not satisfy the requirements?)

5.2.4 Hypothesis Testing with 2SLS (under homoskedasticity: A2SLS3)

Significance test of $\hat{\beta}_j$: asymptotic t-statistics

Test of $H_0 : \mathbf{R\beta} = \mathbf{r}$: Wald Stat. with $\hat{\mathbf{V}} = \hat{\sigma}^2 (\sum_{i=1}^{N} \hat{\mathbf{x}}_i ' \hat{\mathbf{x}}_i)^{-1}$

Wald test stat.: $W_N \equiv (\mathbf{R}\hat{\boldsymbol{\beta}}_N - \mathbf{r})'[\mathbf{R}(\hat{\mathbf{V}}_N / N)\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}_N - \mathbf{r})$

Under H0, $W_N \xrightarrow{d} \chi^2_Q$ (Q:# of restrictions)

5.2.5 Heteroskedasticity-Robust Inference for 2SLS

$$A\mathrm{var}(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{V}} = \left(\sum_{i=1}^{N} \hat{\mathbf{x}}_i ' \hat{\mathbf{x}}_i\right)^{-1} \left(\sum_{i=1}^{N} \hat{u}_i^2 \hat{\mathbf{x}}_i ' \hat{\mathbf{x}}_i\right) \left(\sum_{i=1}^{N} \hat{\mathbf{x}}_i ' \hat{\mathbf{x}}_i\right)^{-1}$$

5.2.6 Potential Pitfalls with 2SLS：Is IV always better than OLS？

1．IV estimator is consistent but biased as long as at least one endogenous variable is in the model

2. If (candidates of) Ivs are not exogenous and correlate only weekly with endogenous variable (weak instrument), bias can be enormous

$Ex. y = \beta_0 + \beta_1 x_1 + u$, x_1: endogenous, z_1: IV for x_1

If $Cov(z_1, u) \neq 0$, $p\lim \hat{\beta}_1 = \beta_1 + Cov(z_1, u) / Cov(z_1, x_1)$

If $Cov(z_1, x_1)$ is close to zero, bias can be close to infinity

In such a case, the OLS bias is smaller than the IVbias

$(p\lim \hat{\beta}_1 = \beta_1 + (\sigma_u / \sigma_{x_1}) Corr(x_1, u))$

Is IV always better than OLS？(cont.)

3．standard error of 2SLS is larger than OLS

IV is not always better than OLS. In the actual estimation, we should test endogeneity (see Section 6.2.1) to determine which method you use

5.3 IV Solutions to the Omitted Variables and Measurement Error Problems

5.3.1 Leaving the Omitted Factors in the Error Term

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \gamma q + v, \quad E(v \mid \mathbf{x}, q) = 0$

In order to estimate $\boldsymbol{\beta}$ consistently with $u = \gamma q + v$, we use 2SLS with IVs s.t.

1．not included in the population model

2．uncorrelated with $u = \gamma q + v$

3．correlated well to endogenous variables with $Cov(x_j, q) \neq 0$

Note: opposite to the proxy variable method

Solution to CEV

The population model : $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \gamma q + v$
observable : $q_1 = q + a_1, Cov(q, a_1) = 0$
Plugging the observable into the population model:
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \gamma q_1 + v - \gamma a_1$
OLS is not consistent （$\because Cov(q_1, v - \gamma a_1) \neq 0$）

Another observable:$q_2 = q + a_2, (Cov(q, a_2) = 0, Cov(a_1, a_2) = 0)$
We can use q_2 as an instrument for q_1

We can estimate consistently the population model if at least two observations for the same variable are available and the measurement errors are uncorrelated each other

Ex. true value educational attainment
observation 1 : answer by the respondent
observation 2 : answer by the spouse

5.3.2 Solutions Using Indicators of the Unobservables

Proxy Variable Solution : use proxy z_1 such that (LP) $q = \theta_0 + \theta_1 z_1 + r_1$ where r_1 is uncorrelated with z_1 (by definition of LP) and uncorrelated with explanatory variables.

Multiple Indicator Solution : consider *indicator* with the following property
LP : $q_1 = \delta_0 + \delta_1 q + a_1$ where $Cov(a_1, q) = 0$ (by definition of LP), $Cov(a_1, \mathbf{x}) = \mathbf{0}$
(redundant in the structural equation)

If $\delta_1 \neq 0$, $q = -(\delta_0/\delta_1) + (1/\delta_1)q_1 - (1/\delta_1)a_1$. Since $Cov(a_1, q_1) \neq 0$,
proxy variable solution is not consistent

5.3.2 Solutions Using Indicators of the Unobservables (cont.)

Another indicator:$q_2 = \rho_0 + \rho_1 q + a_2$ with $\rho_1 \neq 0$.
Assume $Cov(a_1, a_2) = 0$ (indicators correlate each other only through the dependency on q)

Plugging $q = -(\delta_0/\delta_1) + (1/\delta_1)q_1 - (1/\delta_1)a_1$ in $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \gamma q + v$,
$y = \alpha_0 + \mathbf{x\beta} + \gamma_1 q_1 + (v - \gamma_1 a_1)$

q_1 is correlated to q_2, but q_2 is uncorrelated to $v - \gamma_1 a_1$ （since $q_2$is redundant, uncorrelates to v.
Moreover, from $Cov(a_1, a_2) = 0$ and $Cov(a_1, q) = 0, q_2$ is uncorrelated with $\gamma_1 a_1$）
Hence, q_2 is an instrument for q_1

PS 5

- **Reproduce the estimation results in Example 5.3 with Stata (hand in both the output and the DO files)**
- **5.3**
- **5.4**
- **5.12**